

# BASIC TOOL MASTERY

## JUPYTER/PYTHON

BY  
WURAOLA OYEWUSI

# WHAT IS A NOTEBOOK?

A **notebook interface** (also called a **computational notebook**) is a virtual notebook environment used for literate programming. It pairs the functionality of word processing software with both the shell and kernel of that notebook's programming language. Millions of people use notebooks interfaces to analyze data for science, journalism, and education. Wikipedia

Computational notebooks combine code, visualizations, and text in a single document. Researchers, data analysts, and even journalists are rapidly adopting this new medium.

Adam Rule et al, 2018

# Short History of Computational Notebooks

First Computational Notebook By Mathematica (1988)

Maple (1989)

Sage

Jupyter Notebooks are a spin-off project from the IPython project, which used to have an IPython Notebook project itself. The name, Jupyter, comes from the core supported programming languages that it supports: Julia, Python, and R

# What makes Notebooks different from other programming environment

In the Python REPL, only one command can be typed at a time, and only one line of output is shown at a time. In a **.py** file, the entire file is run at one time, line by line. The output of the entire file is produced all at once. Markdown **.md** files contain text in markdown format, but that text is not rendered. In a Jupyter notebook, chunks of code one line or many lines long can be run individually and in any order without running all of the code in the Jupyter notebook. Jupyter notebooks render the markdown sections and display rich text with headings, formatting, and images. Jupyter notebooks contain three types of cells: *code cells*, *output cells*, and *markdown cells*.

- Code cells: Lines of Python code are run in code cells.
- Output cells: The output from running the code cells is also shown in output cells. Charts, plots, command line output, and images can all be shown in Jupyter notebooks as well.
- Markdown cells: Contain text-like descriptions of what will happens in subsequent code cells. Markdown cells can also contain images and links.

# Common Types of Computational Notebooks

JUPYTER

Google Colaboratory Notebook (Support GPU and TPU)

Google Datalab

Kaggle Kernel

IBM Watson Studio (Supports Spark)

Databricks(Free Community Edition)

PolyNote by Netflix

Table of simple text editors and IDE (Integrated Development Environment) which can edit and run Python code

Python Text Editor or IDE	Description
Notepad	Simple text editor - included with Windows
Idle	Included with Python from Python.org
Sublime Text	Full-featured editor with long-time no-cost license
Spyder	IDE included with the Anaconda Distribution of Python
Visual Studio Code	An multi-language open source IDE
PyCharm	Professional Developer-friendly Python IDE

Problemsolvingwithpython.com

# What is Anaconda Distribution

Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment. Package versions are managed by the package management system *conda*. The Anaconda distribution is used by over 15 million users and includes more than 1500 popular data-science packages suitable for Windows, Linux, and MacOS.

wikipedia

```
(base) C:\Users\Wuraola>jupyter notebook
```



```
(base) C:\Users\Wuraola>jupyter notebook
[I 12:57:00.658 NotebookApp] JupyterLab extension loaded from C:\Users\Wuraola\Anaconda3\lib\site-packages\jupyterlab
[I 12:57:00.658 NotebookApp] JupyterLab application directory is C:\Users\Wuraola\Anaconda3\share\jupyter\lab
[I 12:57:00.677 NotebookApp] Serving notebooks from local directory: C:\Users\Wuraola
[I 12:57:00.678 NotebookApp] The Jupyter Notebook is running at:
[I 12:57:00.678 NotebookApp] http://localhost:8888/?token=94394a67b105bd495eba218276ed7ba83bdf4c597f003cff
[I 12:57:00.679 NotebookApp] or http://127.0.0.1:8888/?token=94394a67b105bd495eba218276ed7ba83bdf4c597f003cff
[I 12:57:00.679 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 12:57:00.813 NotebookApp]
```

To access the notebook, open this file in a browser:

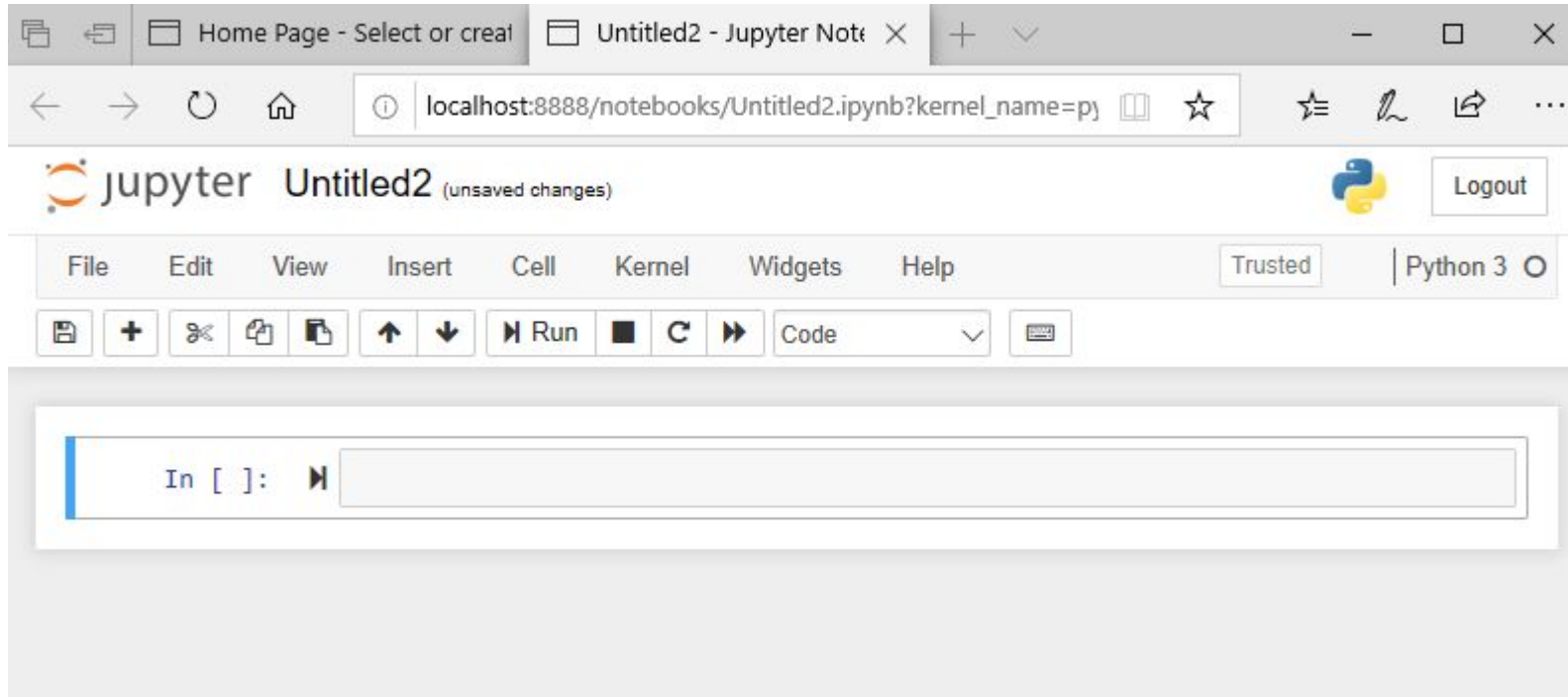
file:///C:/Users/Wuraola/AppData/Roaming/jupyter/runtime/nbserver-15800-open.html

Or copy and paste one of these URLs:

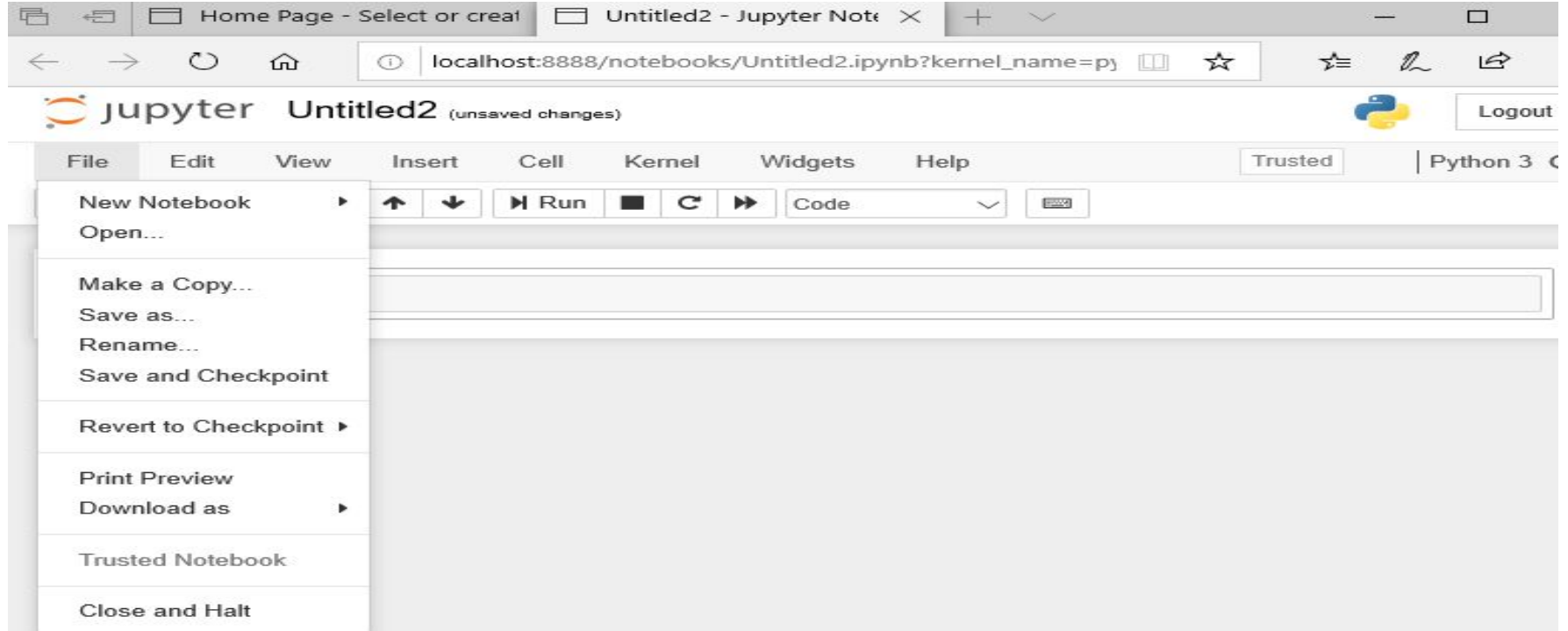
http://localhost:8888/?token=94394a67b105bd495eba218276ed7ba83bdf4c597f003cff

or http://127.0.0.1:8888/?token=94394a67b105bd495eba218276ed7ba83bdf4c597f003cff

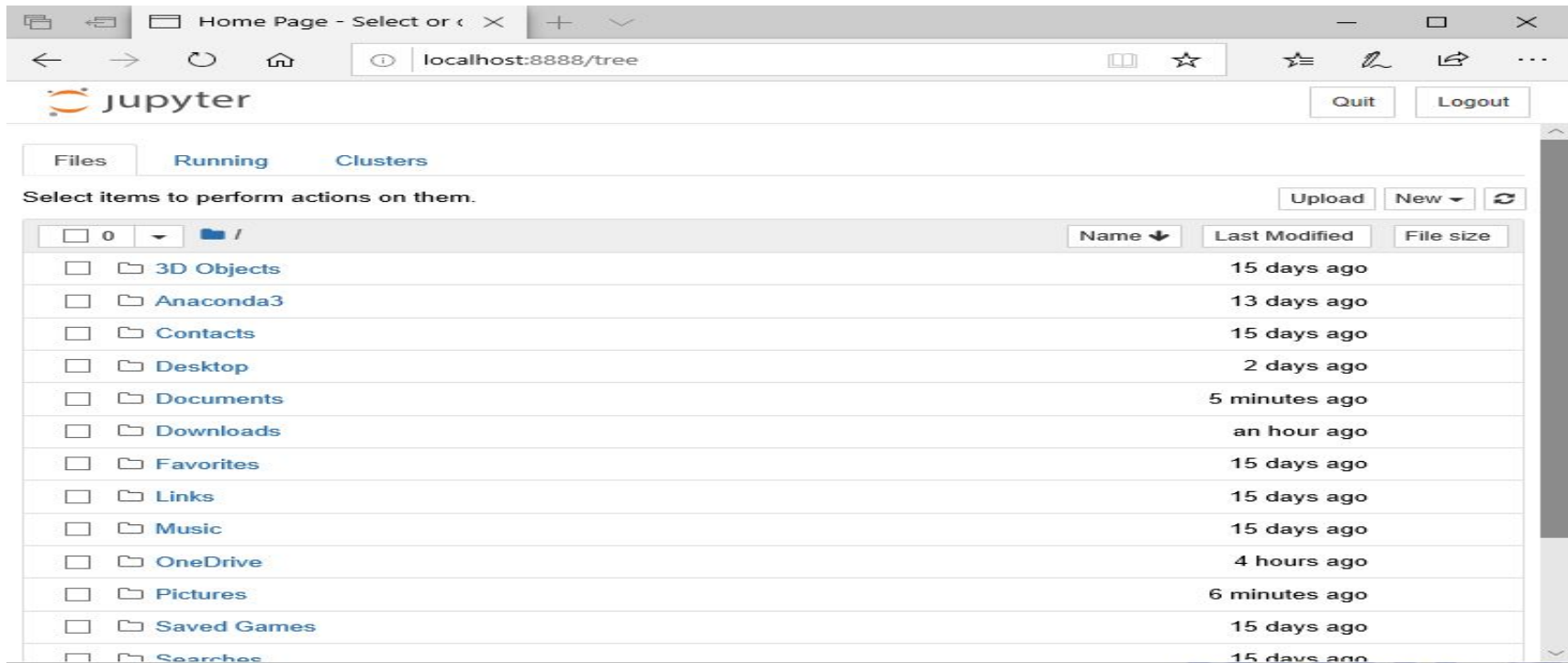
# ANATOMY OF JUPYTER NOTEBOOK



# ANATOMY OF JUPYTER NOTEBOOK



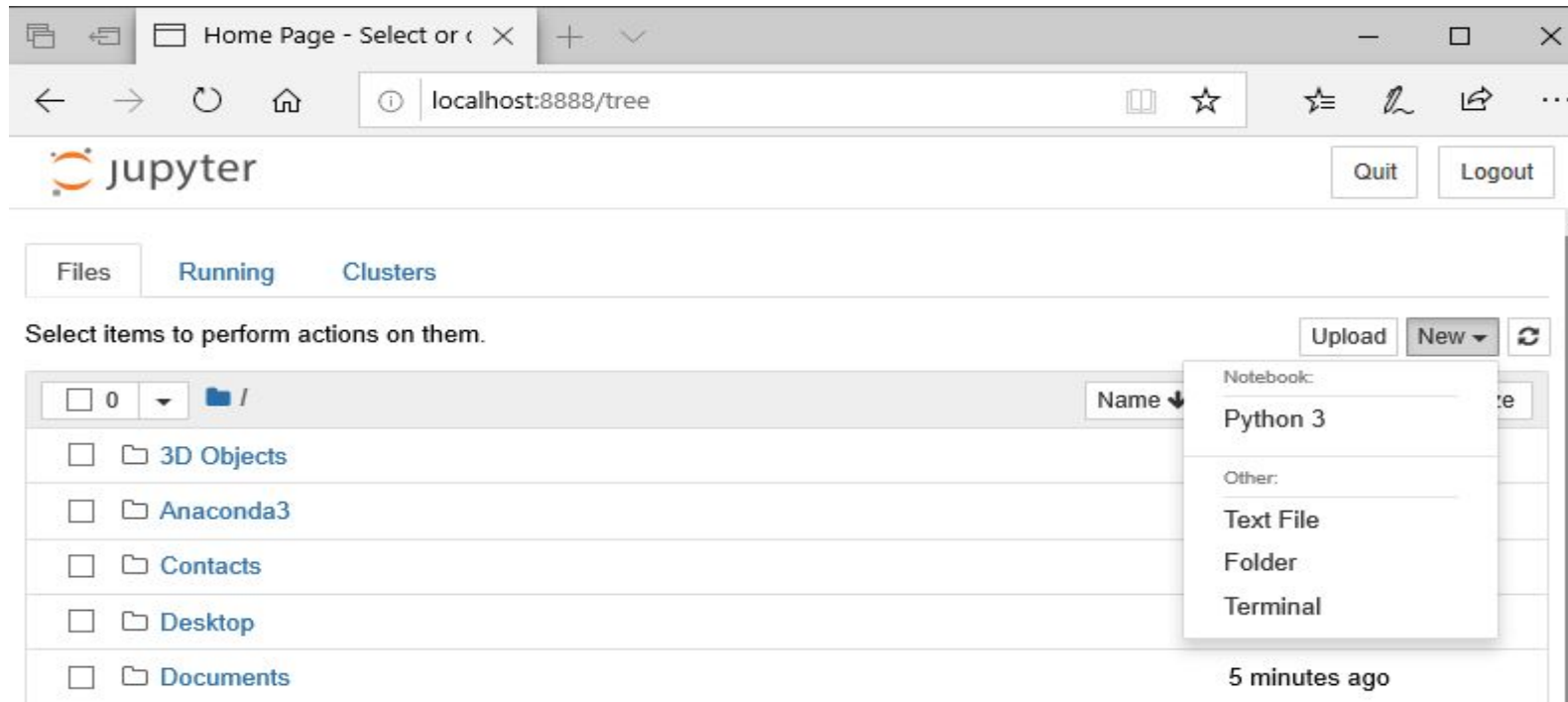
# ANATOMY OF JUPYTER NOTEBOOK



The screenshot displays the Jupyter Notebook web interface in a browser window. The address bar shows the URL `localhost:8888/tree`. The Jupyter logo is visible in the top left, and "Quit" and "Logout" buttons are in the top right. Below the navigation tabs (Files, Running, Clusters), a message states "Select items to perform actions on them." To the right of this message are buttons for "Upload", "New", and a refresh icon. The main area shows a file browser for the root directory (/). It includes a table with columns for "Name", "Last Modified", and "File size". The table lists various system folders with their last modification times.



	Name	Last Modified	File size
<input type="checkbox"/>	0		
<input type="checkbox"/>	/		
<input type="checkbox"/>	3D Objects	15 days ago	
<input type="checkbox"/>	Anaconda3	13 days ago	
<input type="checkbox"/>	Contacts	15 days ago	
<input type="checkbox"/>	Desktop	2 days ago	
<input type="checkbox"/>	Documents	5 minutes ago	
<input type="checkbox"/>	Downloads	an hour ago	
<input type="checkbox"/>	Favorites	15 days ago	
<input type="checkbox"/>	Links	15 days ago	
<input type="checkbox"/>	Music	15 days ago	
<input type="checkbox"/>	OneDrive	4 hours ago	
<input type="checkbox"/>	Pictures	6 minutes ago	
<input type="checkbox"/>	Saved Games	15 days ago	
<input type="checkbox"/>	Searches	15 days ago	

# ANATOMY OF JUPYTER NOTEBOOK



# When your .ipynb is not rendering

nbviewer.jupyter.org

← → ↻  nbviewer.jupyter.org 



[JUPYTER](#) [FAQ](#)

## nbviewer

A simple way to share Jupyter Notebooks

Enter the location of a Jupyter Notebook to have it rendered here:

URL | GitHub username | GitHub username/repo | Gist ID

Go!

---

Programming Languages

# Typical Workflow Notebooks

Define Dataset

Clean and Prepare Dataset

Data Modelling

Interpret the Data

# Best Practices for Writing Notebooks

One notebook, one Focus

Push Codes in modules

Use Speaking Variables and tidy up your Code

Label Diagrams



# What is Python

Python is an easy to learn, powerful programming language. It has efficient high-level data structures and a simple but effective approach to object-oriented programming. Python's elegant syntax and dynamic typing, together with its interpreted nature, make it an ideal language for scripting and rapid application development in many areas on most platforms.

# Python Data Types

A variable's data type is created dynamically, without the need to explicitly define a data type when the variable is created,

Variables in Python are defined with the assignment operator, the equals sign =

- Numeric Data Types E,g Integers,Floating Point Numbers,Complex
- Boolean Data Type E.g True , False
- Strings

# Python Data structures

Data structures are *structures* which can hold some *data* together. In other words, they are used to store a collection of related data.

There are four built-in data structures in Python.

- List : Contains ordered collection of items, separated by commas accessible by indexing, enclosed in square brackets' `[]`
- Tuple : Immutable collection of items enclosed in `()`
- Dictionary: Dictionaries are made up of key: value pairs, Pairs of keys and values are specified in a dictionary by using the notation `d = {key1 : value1, key2 : value2 }`
- Set : Sets are *unordered* collections of simple objects. These are used when the existence of an object in a collection is more important than the order or how many times it occurs.

# Most Popular Python Libraries Used in Data Science

Library	Use
pandas	Data Manipulation & aggregation
numpy	Scientific & technical computing
matplotlib,bokeh	visualization
scikit-learn	Machine learning
tensorflow,pytorch	Deep Learning
nltk,spaCy,gensim	Text processing
beautifulsoup	Web scraping

# REFERENCES

Adam Rule, Aurélien Tabard, James Hollan. Exploration and Explanation in Computational Notebooks. ACM CHI Conference on Human Factors in Computing Systems, Apr 2018, Montréal, Canada. pp.1-12, ff10.1145/3173574.3173606ff. fhal-01676633f

<https://www.kdnuggets.com/2018/11/best-practices-notebooks-data-science.html> by Armin Wasicek

<https://mode.com/blog/data-science-notebook-love-story> by David Wallace

<https://realpython.com/jupyter-notebook-introduction/>

[https://en.wikipedia.org/wiki/Notebook\\_interface](https://en.wikipedia.org/wiki/Notebook_interface)

<https://problemsolvingwithpython.com/02-Jupyter-Notebooks/02.01-What-is-a-Jupyter-Notebook/>

[https://en.wikipedia.org/wiki/Anaconda\\_\(Python\\_distribution\)](https://en.wikipedia.org/wiki/Anaconda_(Python_distribution))

<https://docs.python.org/3.8/tutorial/>

[https://python.swaroopch.com/about\\_python.html](https://python.swaroopch.com/about_python.html)

<https://www.datacamp.com/community/blog/ipython-jupyter>

[How to work with Github in Google Colaboratory Notebook](#)

<https://www.dataquest.io/blog/advanced-jupyter-notebooks-tutorial/>